DOCUMENT RESUME

ED 087 785 TM 003 389

AUTHOR Thomas, David B.
TITLE An APL/1500 Procedure for Generating Random Samples from Trivariate Normal Populations.
PUB DATE 72
NOTE 14p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Computer Assisted Instruction; *Computer Programs; Learning; Measurement Techniques; Models; Research; Sampling; *Simulation; *Teachers; *Training

ABSTRACT

The uses of simulation as a technique for instruction and research have been expanding over the past ten years. As an aid to educational practitioners, the sampling procedures associated with trivariate normal populations are described along with a description of an APL/1500 computer program for this level of multivariate sampling in educational research. A sample use of the program is also given. (Author)

ED 087785

TM 003 389

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

AN APL/1500 PROCEDURE FOR GENERATING RANDOM SAMPLES FROM TRIVARIATE NORMAL POPULATIONS

David B. Thomas

1972

Florida State University

ABSTRACT

The uses of simulation as a technique for instruction and research have been expanding over the past ten years. As an aid to educational practitioners, the sampling procedures associated with trivariate normal populations are described along with a description of an APL/1500 computer program for this level of multivariate sampling in educational research. A sample use of the program is also given.

TABLE OF CONTENTS

ABSTRACT ii

Introduction 1

Sampling Procedure 3

APL/1500 Functions 7

Example Use 8

References 10

AN APL/1500 PROCEDURE FOR GENERATING RANDOM SAMPLES FROM TRIVARIATE NORMAL POPULATIONS

David B. Thomas

Introduction

The uses of simulation as a technique for instruction and research have been expanding over the past ten years. Simulation of psychological theories and the development of "learning machines" have been described by Feigenbaum and Feldman (1963) and others (Green, 1969; Snyder, 1968). Instructional simulations have been suggested as supplements to expository instructional techniques (Raser, 1969). In these instructional applications a formal process from physics, chemistry, or mathematics is modeled and presented to the student as an exercise at a cathode ray tube terminal (Kromhout, Hansen, & Schwarz, 1970). Social science instructional applications are described by Boocock and Schild (1968) and by Raser (1969). In these uses, the social science process may be modeled and presented as a learning game in which the student may act out various roles according to the rules of the game. The use of simulation in teacher training programs is described by Smith and Smith (1966).

At the Florida State University Center for Computer-Assisted Instruction, both the Coursewriter II (IBM, 1968) and APL (Iverson, 1966) programming languages are employed for research in instruction. The

FSU-APL system is implemented on an IBM 1800-based 1500 system which supports both cathode ray tube and teletype-typewriter terminals. The APL/1500 system (Krueger and McMurchie, 1968) permits easy implementation of simulation models based on complex mathematical algorithms. When coupled with the interactive nature of APL, the system provides a simulation capability which may be utilized both for research on instruction and as a means of instruction in mathematics, physics, chemistry, library science, and the social sciences.

An investigation of an innovative teacher training program, with one purpose being the implementation and simulation of a training model on a computer system, is described by King (1970). In this investigation, a methodology was required whereby a normal distribution with a specified mean and variance could be randomly sampled. A method was available which allowed for sampling from these populations. The APL function available employed the method reported by Muller (1959), modified so as to reduce the amount of computer space necessary for execution. Using the function, independent samples could easily be drawn from any normal population with a user specified mean and variance. This APL program is reported by Lippert, et al. (1970) as FSU APL/1500 function FSNU12.

The APL function (FSNU12) was adequate for drawing a series of independent normally distributed random variables. However, as is the case with much research in education and psychology, multiple measurements on the same experimental unit or person were desired. These measurements, rather than being independent, are correlated. In order to simulate phenomena in which multiple measurements on the same persons

were present, a method was required which would sample from a specified multivariate normal population. In such a population, persons are represented as points in an n-dimensional space, wherein each of the variables is considered a dimension. (A multivariate normal population was desired because the majority of statistical measures in education and psychology require the assumption of normality (Morrison, 1967)).

In the multivariate normal population the random variables are related to one another through the variance-covariance matrix Σ. This matrix, in the case of standardized variables, specifies the intercorrelations among the variables. By specifying Σ, one can then interrelate the variables (i.e., the measures on persons) as desired.

The simulation of the teacher training model used by King (1970) required three measures on each person, so that it was necessary to draw random samples from a trivariate normal population. Tocher (1963) presented a method for selecting samples from the bivariate normal population when independently distributed N(0,1) random variables were available. The remainder of this paper describes the rationale for extending Tocher's method, and presents the equations used in the APL/1500 functions to sample from the trivariate normal populations. Examples of the function's use are also given.

Sampling Procedure

Suppose that the trivariate normal population of interest has the following variance-covariance matrix, Σ, and mean vector μ

$$\Sigma = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{21} & 1 & \rho_{23} \\ \rho_{31} & \rho_{32} & 1 \end{bmatrix}$$

$$\mu = [0 \quad 0 \quad 0]$$

in which the ρ_{ij} are the population values of the correlations between variables i and j. If Y_1, Y_2, and Y_3 are three normally and independently distributed random variables, each N(0, 1), then the trivariate normal random variables X_1, X_2, X_3 may be represented as functions of the Y_i as follows:

(1) $X_1 = Y_1$

(2) $X_2 = a_1Y_1 + a_2Y_2$

(3) $X_3 = a_3Y_1 + a_4Y_2 + a_5Y_3$

The task is to compute the constants a_i such that a specific variance-covariance matrix, Σ, of the variables X_1, X_2, X_3 results.

For any two normally and independently distributed random variables Y_i and Y_j, it is known (Anderson, 1958; Hogg and Craig, 1970) that

(4) $E[Y_iY_j] = 0.$

For correlated standardized random variables Y_i and Y_j

(5) $E[Y_iY_j] = \rho_{ij}.$

From the requirement that the Y's be distributed normally with mean zero and variance 1, it is know that

(6) $E[Y_i^2] = 1$

From these properties, the values of the weights a_i can be derived as follows:

$\rho_{12} = E[X_1X_2]$ from (5)

$= E[Y_1(a_1Y_1 + a_2Y_2)]$ from (1) and (2)

So that

(7) $\rho_{12} = a_1$ by expansion and from (4) and (6)

Now to derive a_2,

$$1 = E[X_2^2] \qquad \text{from (6)}$$

$$= E[(a_1Y_1 + a_2Y_2)^2] \qquad \text{from (2)}$$

and in a similar manner, we show:

$$(8) \qquad a_2 = \sqrt{1-\rho_{12}^2}$$

To derive a_3, we set:

$$\rho_{13} = E[X_1X_3] \qquad \text{from (5)}$$

$$= E[Y_1(a_3Y_1 + a_4Y_2 + a_5Y_3)] \qquad \text{from (1) and (3)}$$

from which

$$(9) \qquad \rho_{13} = a_3$$

To derive a_4, set

$$\rho_{23} = E[X_2X_3] \qquad \text{from (5)}$$

$$= E[(a_1Y_1 + a_2Y_2)\,(a_3Y_1 + a_4Y_2 + a_5Y_3)], \qquad \text{from (2) and (3)}$$

and by procedures similar to those above, we find

$$\rho_{23} = a_1a_3 + a_2a_4$$

$$= \rho_{12}\rho_{13} + a_4\sqrt{1-\rho_{12}^2} \qquad \text{from (7), (8), and (9)}$$

from which

$$(10) \qquad a_4 = \frac{\rho_{23}-\rho_{12}\rho_{13}}{\sqrt{1-\rho_{12}^2}}$$

finally, to computer a_5:

$$1 = E[X_3^2] \qquad \text{from (6)}$$

$$= E[(a_3Y_1 + a_4Y_2 + a_5Y_3)^2] \qquad \text{from (3)}$$

by expanding and substituting we can show that

$$1 = \rho_{13}^2 + \frac{(\rho_{23}-\rho_{12}\rho_{13})^2}{1-\rho_{12}^2} + a_5^2$$

from which

(11) $$a_5 = \sqrt{1-\rho_{13}^2-\frac{(\rho_{23}-\rho_{12}\rho_{23})^2}{1-\rho_{12}^2}}$$

To summarize,

(7) $$a_1 = \rho_{12}$$

(8) $$a_2 = \sqrt{1-\rho_{12}^2}$$

(9) $$a_3 = \rho_{13}$$

(10) $$a_4 = \frac{\rho_{23}-\rho_{12}\rho_{13}}{\sqrt{1-\rho_{12}^2}}$$

(11) $$a_5 = \sqrt{1-\rho_{13}^2-\frac{(\rho_{23}-\rho_{12}\rho_{23})^2}{1-\rho_{12}^2}}$$

From the foregoing, the following restrictions emerge:

(1) $(\rho_{12}^2 \neq 1)$ in order that a_4 may be computed (has finite bounds)

(2) $\rho_{12}^2+\rho_{13}^2+\rho_{23}^2 \leq 1+2\rho_{12}\rho_{13}\rho_{23}$ (in order that a_5 remain real).

The X_i may now be computed from the already selected Y_i. The X_i will then constitute a random sample from the trivariate normal population with the variance-covariance matrix Σ and mean vector μ.

APL 1500 Functions

The function M employs a modification of the Box-Muller method for generating standardized normal random deviates (see Muller, 1959). These are then transformed into random variables from a normal distribution with given mean μ and variance σ^2 before being output. See Lippert, et al. (1970) for a description of the use of this function independently of the current application (FSNU11). A listing of the function M follows.

```
      ∇M[⎕]∇
    ∇ R←N M M
[1]   R←(N⍴M[;1])+(N⍴M[;2]*0.5)×(100○×(?N⍴R)÷R)×(2×●R÷?(N←N,1⍴⍴M
      )⍴R←8388608)*0.5
    ∇
```

The function X computes the constants a_i and derives the trivariate sample points X_1, X_2, X_3 by employing the function M to generate independent N(0,1) variables which are manipulated according to the procedure given in equations (1)-(3) from the previous section. The explicit result of X is an r x 3 matrix of data in which each row corresponds to the standardized coordinates of the sample point in three dimensional space; r sample points are generated. Other straightforward measures are required to transform these coordinates into data with given means and variances (i.e., these data are standardized normal variates). A listing of this function follows:

```
      ∇X[⎕]∇
    ∇ R←N X C;A;B;D
[1]   →((+/C*2)>1+2××/C)⍴6
[2]   A←C[1],B,C[2],D,(1-(C[2]*2)+(D←(C[3]-×/C[1 2])÷B←(1-C[1]*
      2)*0.5)*2)*0.5
[3]   R[;3]←(R←N M 3 2 ⍴ 0 1)+.×¯3↑A
[4]   R[;2]←R[;1 2]+.×2↑A
[5]   →0
[6]   R←⍳0,0⍴⎕←'CORRELATIONS OUTSIDE DOMAIN OF FUNCTION.'
    ∇
```

Example Use

Suppose that a sample of size 20 drawn from a trivariate population was required with the following covariance matrix:

$$\Sigma = \begin{bmatrix} 1.0 & .5 & .1 \\ .5 & 1.0 & -.6 \\ .1 & -.6 & 1.0 \end{bmatrix}$$

The following APL expression would generate the required sample:

```
20 X .5 .1 ¯.6
```

Sample data are given below:

¯0.7454	¯2.019	0.1944
¯1.796	¯0.8288	¯0.01032
¯0.943	0.471	¯1.363
¯0.4799	¯0.4117	¯1.041
¯1.236	0.441	¯0.6187
1.23	¯0.3218	0.3429
0.3484	¯1.037	0.521
0.3126	1.79	¯1.117
1.023	¯0.004035	¯0.2626
0.7774	1.033	¯0.02037
¯0.4948	0.01833	¯0.2669
1.648	1.411	¯0.4178
¯0.1092	¯0.2843	0.3142
¯0.1421	¯0.1251	¯0.3569
1.118	¯0.05234	1.663
¯0.9683	¯0.5068	¯0.2334
¯0.2228	¯1.664	2.582
¯0.5176	¯0.9099	0.4406
¯0.09321	¯0.06726	0.464
¯0.7093	¯1.898	1.597

The sample correlation matrix for these data is:

$$R = \begin{bmatrix} 1.00 & .47 & .09 \\ .47 & 1.00 & -.62 \\ .09 & -.62 & 1.00 \end{bmatrix}$$

These values are reasonable given the size of the sample and the usual sampling errors.

It is hoped that these procedures and their logical extensions will prove useful to behavioral scientists engaged in training research using simulation as a tool. At the present time, educational usages of this nature are limited. These functions were employed in an investigation of a teacher training model (King, 1970). Hopefully, publication of these functions will stimulate additional applications of the technique.

REFERENCES

Anderson, T. W. Introduction to multivariate statistical analysis. New York: Wiley, 1958.

Boocock, S. S. and E. O. Schild (Eds.), Simulation games in learning. Beverly Hills, Calif.: Sage Publications, Inc., 1968.

Feigenbaum, E. A., and J. Feldman (Eds.) Computers and thought. San Francisco: McGraw-Hill, 1963.

Green, B. F., Jr. The computer conquest of psychology, in Readings in Psychology Today. Del Mar, Calif., CRM Books, 1969.

Hogg, R. V., and A. T. Craig. Introduction to mathematical statistics. London: Macmillan, 1970.

International Business Machines Corporation. 1500 coursewriter II authors guide. San Jose: IBM Corp., 1968.

Iverson, K. E. Elementary functions: An algorithmic treatment. Chicago: Science Research Associates, 1966.

King, A. D. An application of simulation techniques to an innovative teacher training program. Unpublished doctoral dissertation, Florida State University, 1970.

Kromhout, O. M., D. N. Hansen, and G. Schwarz. Conference on computers in undergraduate science education: A computer-assisted and managed course in physical sciences. Paper presented at Proceedings of the Congress on Computers in Undergraduate Science Education, Illinois Institute of Technology, August, 1970.

Krueger, S., and T. McMurchie. APL/1500 Users Guide. Chicago: IBM Corporation, 1968.

Lippert, H. T., E. V. Harris, D. B. Thomas, and T. D. McMurchie. Manual of APL/1500 functions. Systems Memo 3, Florida State University, Tallahassee, 1970.

Morrison, D. F. Multivariate statistical methods. New York: McGraw-Hill, 1967.

Müller, M. E. A comparison of methods of generating normal deviates on digital computers. J. Assoc. Comp. Mach., 1959, 6, 376-383.

Raser, J. R. Simulation and society: An exploration of scientific gaming. Boston, Allyn and Bacon, Inc., 1969.

Smith, K. V., and M. F. Smith Cybernetic principles of learning and educational design. New York: Holt, Rinehart and Winston, 1966.

Snyder, H. I. Contemporary educational psychology: Some models applied to the school setting. New York: Wiley, 1968.

Tocher, K. D. The art of simulation. Princeton, N.J.: Van Nostrand, 1963.